REMARKS

The foregoing amendments are made to insert the sequence identification numbers into the specification. No new matter has been added.

Respectfully submitted,

Date: Jan 30, 2002 By: D Pasternak
Dahna S. Pasternak
Registration No. 41,411

ROBINS & PASTERNAK LLP
90 Middlefield Road, Suite 200
Menlo Park, CA 94025
Telephone: 650-325-7812
Facsimile: 650-325-7823

Version with markings to show changes made

In the specification:

Paragraph beginning on page 1, line 11 has been amended as follows:

Zinc finger proteins (ZFPs)are proteins that can bind to DNA in a sequence-specific manner. Zinc fingers were first identified in the transcription factor TFIIIA from the oocytes of the African clawed toad, *Xenopus laevis*. An exemplary motif characterizing one class of these protein (C_2H_2 class) is -Cys-$(X)_{2-4}$-Cys-$(X)_{12}$-His-$(X)_{3-5}$-His (where X is any amino acid) (SEQ ID NO:1). A single finger domain is about 30 amino acids in length, and several structural studies have demonstrated that it contains an alpha helix containing the two invariant histidine residues and two invariant cysteine residues in a beta turn co-ordinated through zinc. To date, over 10,000 zinc finger sequences have been identified in several thousand known or putative transcription factors. Zinc finger domains are involved not only in DNA-recognition, but also in RNA binding and in protein-protein binding. Current estimates are that this class of molecules will constitute about 2% of all human genes.

Paragraph beginning on page 5, line 3 has been amended as follows:

Fig. 3 (A) Sketch showing key segments of the phagemid. (B) Expected arrangement of fusion proteins at the target DNA. Phage displaying two copies of a dimerizing peptide-Zif12 fusion can form stable complexes with the biotinylated target DNA site, which contains an inverted repeat of the Zif12-binding site. The phage-DNA complexes are captured by streptavidin coupled to a solid support, and phage that bind less tightly are washed away. (C) The DNA site (SEQ ID NO:23) used for affinity selection of phage, with the two juxtaposed Zif12-binding sites in bold.

Paragraph beginning on page 5, line 10 has been amended as follows:

Fig. 4 (A) Sequences of peptide extensions (SEQ ID NOS:24-29) isolated from the initial selection. Numbers in parentheses give the frequency of occurrences among the 45 clones sequenced. The clones for peptide 2 included a Glu-21-to-Asp mutation in the zinc finger region that may have been partially responsible for the affinity of this peptide. (B) Gel mobility-shift assays using purified fusion peptides 1, 3, and 5. Protein (2.5 μM, 250 nM, 25 nM, 250 pM, and no protein) was incubated with DNA containing either an inverted repeat of Zif12 sites or a single Zif12 site and then electrophoresed through native polyacrylamide gels. The reduced mobility of the inverted repeat probe in the presence of protein indicates the formation of protein-DNA complexes. Similar results were obtained with fusion peptides 2 and 6. Binding of peptide 4 depend on disulfide bond formation.

Paragraph beginning on page 6, line 1 has been amended as follows:

Fig. 6. Evolution of peptides 1 and 5 by sequential block reoptimization (SEQ ID NOS:30-79). The sequences selected from each reoptimization step are shown in bold, with the number of selection and amplification cycles given in parentheses. Sequences roughly matching the consensus that were used in later steps have been boxed. In some cases, such as in reoptimization step 3 for peptide 1 and reoptimization step 1 for peptide 5, we isolated clones that carried spurious mutations at a nondegenerate position of the peptide extension. In addition, the E21D mutation in the zinc finger region (which also was seen in the original peptide 2 sequence) arose several times; this mutation may stabilize complex formation by improving contacts at the protein-DNA interface. Note: some confusion was caused by this E21D mutation, which occurred in the first reoptimization step for peptide 1, but was discovered only after reoptimization step 2. Thus, the "consensus" sequence from reoptimization step 1 (VPKQR (SEQ ID NO:53)), chosen after selection-amplification cycle 9, had a glutamine that did not occur in sequences isolated after cycle 6. To double-check this position of peptide 1, it was

randomized again during reoptimization step 3. The corresponding position was allowed to vary as Q, M, I, or L, along with the complete randomization of the third block. The reselections showed that methionine or leucine is preferred at this position.

Paragraph beginning on page 6, line 18 has been amended as follows:

Fig. 7: Comparison of the contact surfaces of various zinc fingers. Zif268 finger 1 (SEQ ID NO:80), GLI finger 1 (SEQ ID NO:81), GLI finger 2 (SEQ ID NO:82), and SW15 finger 1 (SEQ ID NO:83) are shown. (A) Stereoview of a superposition of the four fingers . Because of the different lengths of the fingers, the superposition aligned C atoms of residues 104-114 and 116-127 of Zif finger 1 with those of residues 3-13 and 15-26 of GLI finger 1; residues 105-114 and 116-130 of Zif finger 1 with 37-46 and 51-65 of GLI finger 2; and residues 104-130 of Zif finger 1 with 31-57 of SWI5 finger 1. Side chains on each finger that are involved in hydrophobic contacts at the corresponding protein-protein interface have been rendered in stick representation. (Coordinates for GLI and SW15 are from Pavletich *Science* **261**, 1701 (1993); Dutnall, *Structure* **4**, 599 (1996).) (B) Alignment of the sequences of the fingers, with interacting residues boxed. The portion of Zif268 shown here corresponds to residues 104 to 130 in the crystal structure.

Paragraph beginning on page 9, line 19 has been amended as follows:

A D-able subsite within a target site has the motif 5'NNGK3' (SEQ ID NO:2). A target site containing one or more such motifs is sometimes described as a D-able target site. A zinc finger appropriately designed to bind to a D-able subsite is sometimes referred to as a D-able finger. Likewise a zinc finger protein containing at least one finger designed or selected to bind to a target site including at least one D-able subsite is sometimes referred to as a D-able zinc finger protein.

Paragraph beginning on page 15, line 16 has been amended as follows:

T G E K P: (SEQ ID NO:3) (Liu et al., 1997, supra.); (G4S)n (SEQ ID NO:4) (Kim et al., *PNAS* 93, 1156-1160 (1996.); GGRRGGGS (SEQ ID NO:5); LRQRDGERP (SEQ ID NO:6); LRQKDGGGSERP (SEQ ID NO:7); LRQKD(G3S)2 ERP (SEQ ID NO:8). Alternatively, flexible linkers can be rationally designed using computer program capable of modeling both DNA-binding sites and the peptides themselves or by phage display methods . In a further variation, noncovalent linkage can be achieved by fusing two zinc finger proteins with domains promoting heterodimer formation of the two zinc finger proteins. For example, one zinc finger protein can be fused with fos and the other with jun (see Barbas et al., WO 95/119431).

Paragraph beginning on page 15, line 24 has been amended as follows:

A component finger of zinc finger protein typically contains about 30 amino acids and has the following motif (N-C) (SEQ ID NO:1):

Cys-$(X)_{2-4}$-Cys-X.X.X.X.X.X.X.X.X.X.X-**His**-$(X)_{3-5}$-His

-1 1 2 3 4 5 6 7

Paragraph beginning on page 15, line 28 has been amended as follows:

The two invariant histidine residues and two invariant cysteine residues in a single beta turn are co-ordinated through zinc (see, e.g., Berg & Shi, *Science* 271, 1081-1085 (1996)). The above motif shows a numbering convention that is standard in the field for the region of a zinc finger conferring binding specificity. The amino acid on the left (N-terminal side) of the first invariant His residues is assigned the number +6, and other amino acids further to the left are assigned successively decreasing numbers. The alpha helix begins at residue 1 and extends to the residue following the second conserved histidine. The entire helix is therefore of variable length, between 11 and 13 residues. The process of designing or selecting a nonnaturally occurring or variant ZFP typically

starts with a natural ZFP as a source of framework residues. The process of design or selection serves to define nonconserved positions (i.e., positions -1 to +6) so as to confer a desired binding specificity. One suitable ZFP is the DNA binding domain of the mouse transcription factor Zif268. The DNA binding domain of this protein has the amino acid sequence:

YACPVESCDRRFSRSDELTRHIRIHTGQKP (F1) (SEQ ID NO:9)

FQCRICMRNFSRSDHLTTHIRTHTGEKP (F2) (SEQ ID NO:10)

FACDICGRKFARSDERKRHTKIHLRQK (F3) (SEQ ID NO:11)

and binds to a target 5' GCG TGG GCG 3'.

Paragraph beginning on page 16, line 12 has been amended as follows:

Another suitable natural zinc finger protein as a source of framework residues is Sp-1. The Sp-1 sequence used for construction of zinc finger proteins corresponds to amino acids 531 to 624 in the Sp-1 transcription factor. This sequence is 94 amino acids in length. The amino acid sequence of Sp-1 is as follows

PGKKKQHICHIQGCGKVYGKTSHLRAHLRWHTGERP

FMCTWSYCGKRFTRSDELQRHKRTHTGEKK

FACPECPKRFMRSDHLSKHIKTHQNKKG (SEQ ID NO:12)

Sp-1 binds to a target site 5'GGG GCG GGG3'.

Paragraph beginning on page 16, line 20 has been amended as follows:

An alternate form of Sp-1, an Sp-1 consensus sequence, has the following amino acid sequence:

meklrngsgd

PGKKKQHACPECGKSFSKSSHLRAHQRTHTGERP

YKCPECGKSFSRSDELQRHQRTHTGEKP

YKCPECGKSFSRSDHLSKHQRTHQNKKG (SEQ ID NO:13) (lower case letters are a leader sequence from Shi & Berg, *Chemistry and Biology* 1, 83-89. (1995). The optimal binding sequence for the Sp-1 consensus sequence is 5'GGGGCGGGG3'. Other suitable ZFPs are described below.

Paragraph beginning on page 31, line 14 has been amended as follows:

Labeled DNA probes were generated as follows. for the gel-shift studies shown in Fig. 2B, oligos corresponding to the phage-selection target site (5'GGTTGCAG**TGGGCGCGCCCA**CAGTACTTGAACGTAACG-3' (SEQ ID NO:14) and 5'-CGTTACGTTCAAGTACTG**TGGGCGCGCCCA**CTGC-3' (SEQ ID NO:15), Zif12 sites in bold) or a single-site mutant (bold regions above replaced with the sequences 5'-**TGGGCG**TATGCT-3' (SEQ ID NO:16) and 5'AGCATA**CGCCCA**-3' (SEQ ID NO:17)) were annealed and end-labeled with Klenow. A labeled restriction fragment was used for quantitative studies. The oligos 5'-GGAATTCCTGATCAAGATCTGGTCACGTCCATAGGCTAGGCATGTCAAGGCT GTATG-3' (SEQ ID NO:18) and 5'-GGGATCCACTCGCGAACGCGTCCTTGTAG**TGGGCGCGCCCA**CATACAGCCT TGACAT-3' (SEQ ID NO:19) (Zif12 sites in bold) were annealed, extended by mutually primed extension, and cloned into the *Eco*RI and *Bam*HI sites of pBluescript II SK(+). The probe was prepared by digesting the plasmid with *Eco*RI and *Not*I; labeling the DNA with Klenow, (α -^{32}P)dCTP, and (α -^{32}P)dGTP; and purifying the small fragment by native PAGE.

Paragraph beginning on page 33, line 20 has been amended as follows:

Table 1. Affinities for DNA duplexes containing the inverted repeat site 5'-TGGGCGCGCCCA-3' (SEQ ID NO:20).

Paragraph beginning on page 43, line 4 has been amended as follows:

Table 2: Crystallographic Analysis. The zinc finger fusion peptide (containing an NH3-Met-Glu-Pro leader peptide, the 15 residues of the selected peptide extension (Wang et al. Proc. Natl. Acad. Sci. USA 90, 9568 (1999)), and residues 4 to 60 of Zif268 (Pavletich, *supra*) and was overexpressed, purified and prepared for crystallization essentially as described previously for Zif268 variant peptides in Elrod-Erickson et al. Structure 6, 451 (1998)) and was cocrystallized with a 14-base pair DNA duplex (the self-complementary oligonucleotide 5'-ATGGGCGCGCCCAT-3' (SEQ ID NO:21) was purified as described previously in Klemm et al., Cell 77,21 (1994) and annealed at a high concentration (3 mM in duplex) to favor formation of intermolecular duplexes over intramolecular hairpins. Derivative oliogs contained either 5-iodouracil at position 1 or 5-iodocytosine at position 8) using PEG 4000 as the precipitant. Equal volumes of protein (1.1 mM in dimer) and duplex DNA (1.5 mM) were mixed, and complexes were solubilized with the addition of NaCl to 0.4 M. Crystals were grown in an anaerobic chamber using hanging drop vapor diffusion from drops containing the complex and an equal volume of the well solution (13-20% PEG-4,000/50-150 mM NaCl/10 mM $MgCl_2$/50 mM MES, pH 6.2). The crystals were soaked for approximately 5 minutes in a solution containing equal volumes of well solution and cryoprotectant solution 38% glycerol/20% PEG-4,000/100 mM NaCl/50 mM MES, pH 6.2) and flash-cooled in a stream of cold nitrogen (126 K). The crystals form in spacegroup $P3_1$ with cell dimensions $a = b = 86.2$ Å, $c = 133.0$ Å. Diffraction data (for Native-1 and the derivatives) were first collected on cryocooled crystals using a rotating anode X-ray generator and an R-Ax 8 IV image plate system. Data were processed with the HKL suite (Z. Otwinowski, W. Minor, *Methods Enzym.* **276**, 307 (1997)). The crystals exhibited partial merohedral twinning, and the twin fraction in each crystal (as listed in the Table) was estimated according to the procedure of Yeates (T.O. Yeates, *Methods Enzym.* **276**, 344 (1997)). Data was detwinned using the DETWIN program in the CCP4 suite

(Collaborative Computational Project Number 4, *Acta Crystallogr.* **D50**, 760 (1994)). There are three dimer-DNA complexes per asymmetric unit (In the crystal, DNA duplexes stack end-to-end, although the basepairs at the junctions are rotationally offset so that pseudocontinuous helices are not formed. DNA stacks run along each crystallographic 3_1 screw axis. Thus, there are three stacks of three complexes in each unit cell, and complexes within each stack are crystallographically related. The three crystallographically independent complexes (and the two crystallographically independent halves of each complex) are very similar to one another. However, one of the protein monomers shows poor density for several residues at the N-terminus (Asn 93 to Val 98) and for finger 2, especially in the beta hairpin region. (There are few crystal contacts in these regions, and they therefore may be more mobile.) With these exceptions, residues Asn 93 to Thr 158 of each protein monomer and the entire DNA duplex for each complex are visible in our structure.) Iodine sites in the derivatives were located with SOLVE (T. Terwilliger, J. Berendzen, *Acta Crystallogr.* **D55**, 849 (1999)), and MIR phases (SOLVE located 6 of 6 sites in derivative IdU-2 and 5 of 6 sites in derivative IdC-8. Since there were three copies of each duplex in the asymmetric unit, and because each duplex was expected to have the same distribution of iodine sites, superimposing corresponding sets of heavy atom positions allowed us to predict the remaining heavy atom site and to determine approximate noncrystallographic symmetry (NCS) operators) were improved by solvent flattening and noncrystallographic symmetry (NCS) averaging using DM (Collaborative Computational Project Number 4, *Acta Crystallogr.* **D50**, 760 (1994)). The resulting experimental electron density map was readily interpretable, and a model was built with O (T.A. Jones, J.-Y. Zou, S.W. Cowan, M. Kjeldgaard, *Acta Crystallogr.* **A47**, 110 (1991)). We refined the model to 2.7 Å with X-PLOR (A.T. Brunger, *X-PLOR Version 3.1: A System for X-ray Crystallography and NMR* (Yale University Press, New Haven, CT, 1992)) using the Native-1 data set. As refinement progressed, we relaxed NCS constraints to restraints and then eliminated NCS

restraints altogether, refined grouped B-factors, and applied a bulk solvent correction. We then collected an additional data set (Native-2) at the National Synchrotron Light Source on Beamline X4A (λ =1.0093 Å), and this data was detwinned and merged with the detwinned Native-1 data. Using this higher-resolution data set, we proceeded with further positional refinement and individual, restrained B-factor refinement and also added 319 water molecules to the model. In the final model, 90.9% of the residues lie in the core regions of the Ramachandran plot and the remaining residues occupy additional allowed regions.

Paragraph beginning on page 46, line 5 has been amended as follows:

As observed with the homeodomain heterodimers, hydrophobic interactions dominate the peptide-protein interface in our complex (Fig. 9B). In describing these interactions, we use a numbering scheme that follows the convention used in the wild-type Zif268 structure (residue numbers 104 to 160 in the crystal structure correspond to residues 4 to 60 in the zinc finger sequence, Pavletich and Pabo, Science 252, 809 (1991)), and thus our 15-residue peptide extension is numbered as H89-PMNNLLNYVVPKM-R103 (SEQ ID NO:22) (to indicate that this is the preceding region of the polypeptide chain in our new protein). Near the twofold axis at the center of the site, the side chain of Pro 104 packs against Pro 104' and Tyr 105' (from the other subunit), while Met 102 interacts with Pro 104, Pro 104', and Tyr 105'. In this region, there is also a hydrogen bond between the carbonyl oxygen of Ser 117 and the hydroxyl group of Tyr 105'. Further outward along the peptide, the side chains of Val 99 and Pro 100 form a nonpolar surface that supports the side chain of Tyr 97 (which also interacts with Leu 94). Leu 94, Leu 95, Tyr 97, and Val 99 of the peptide extension contact a number of residues in zinc finger 1 of the other monomer and thus form a key part of the dimer interface. Leu 94 fits into a hydrophobic pocket formed by zinc finger residues Pro 108', Val 109', Ile 126', His 129', and Thr 130. Leu 95 contacts nonpolar groups on the

side chains of Thr 130' and Gln 132'; Tyr 97 touches Pro 108' and Ile 126', and Val 99 interacts with Try 105', Ser 199', Leu 122', Thr 123', and Ile 126'. Along the edges of this extensive, hydrophobic interface, there are several bridging water molecules, but the water-mediated interactions that we observe are not conserved among the crystallographically independent copies of the dimer interface.